When Three Heads Are Better Than Two

Jennifer Wiley (jwiley@uic.edu)
Department of Psychology, 1007 W. Harrison Street (M/C) 285
Chicago, IL 60607 USA

Melinda S. Jensen (jensenm@uic.edu)
Department of Psychology, 1007 W. Harrison Street (M/C) 285
Chicago, IL 60607 USA

Abstract

The present study investigated the effects of both group size and heterogeneity in relevant math skill on collaborative problem solving. Overall, triads demonstrated better reasoning and more effective cryptarithmetic problem solving than dyads and singletons. Importantly, triads outperformed the best individuals in a nominal groups analysis. Triads were also better able than dyads to take advantage of heterogeneity in math skill. The results suggest that some specific advantages of group collaboration may be more likely to be realized in triads than dyads, and that triads may be an optimal group size when critical evaluation and reasoning are required for a task.

What Are The Advantages of Collaboration?

One popular reason why people believe groups should be more effective, flexible and innovative at problem solving is the assumption that each group member brings to the task a slightly different set of task-relevant knowledge and skills. Through discussion, the knowledge and skills of each member can become available for all, giving each member a larger pool of ideas to draw from. Especially if members possess different backgrounds, group problem solving will give people a greater opportunity for novel associations, strategies and operations. Exposure to diverse viewpoints may increase both the quantity and quality of idea generation in a group context (Hoffman, Harburg & Maier, 1962; Maier, 1962; Nemeth, 1986; Paulus & Yang, 2000, Stasser, Stewart, & Wittenbaum, 1995).

There are a few studies in the cognitive science literature that have explored the idea that diversity and collaboration may be important for successful and innovative scientific discovery. For example, in an investigation of several molecular biology laboratories, Dunbar (1997) has reported that the diversity of a group can be very important. When scientists in a laboratory are from diverse backgrounds, they are better able to generate alternative hypotheses and analogies in the face of unexpected findings, which can in turn lead to scientific breakthroughs.

In an attempt to examine this phenomena in an experimental context, Wiley and Jolly (2003) examined collaborative performance on a creative problem solving task where prior knowledge has been shown to be related to fixation and an inability to come to solution (Wiley, 1998). Comparisons of observed versus expected outcomes (based on combinations of individual base rates) indicated that only the mixed knowledge pairs with one high knowledge member and one low knowledge member performed significantly better than would be predicted using base rates. The high knowledge pairs also tended to do better than would be expected, but this did not reach significance. The low knowledge pairs did significantly worse than would be expected based on past performance of individual low knowledge participants. This result shows specific advantages for experts collaborating when their prior knowledge may lead them into fixation on unpromising solution paths, and that experts may sometimes need the assistance of novices in order to be most effective, flexible or innovative in their problem solving.

However, it is also the case that collaboration does not always lead to superior outcomes (Hill, 1982; Taylor, Berry & Block, 1958; Salomon & Globerson, 1989; Steiner, 1972; Stroebe & Diehl, 1984). In another study on collaborative argumentation, we found disadvantages to students working in dyads (Wiley & Bailey, 2006). Instead of acting as critical evaluators of information and pushing the reasoning to a higher level, we found pairs of students engaging in a lot of passive acceptance of each others' statements. This highlights the need to examine the many possible advantages of collaboration, and under which specific contexts hypothetical advantages may be most likely to occur. In the case of the present study, we were interested in both the size of the group and the heterogeneity of math skills within the group as possible determinants of benefits from collaboration.

Two Heads or Three?

The decision to assign students to dyads or triads in the context of the classroom activities is often based on considerations such as how many computers or materials sets are available, how many learning products (papers, worksheets or projects) the teacher wants to grade, how desks are arranged, and so forth. However, even in the absence of the practical limitations of the classroom, it is not clear which size unit should be expected to do better theoretically. Dyads have a number of possible advantages: each student has more of an opportunity to participate, and there are fewer group members to distract them from their own thinking (Dugosh, Paulus, Roland & Yang, 2000). In addition, there is less of a chance that students will "free-ride" or "loaf" in a group of two, and the student is more likely to feel more highly invested in the product or activity (Stroebe & Diehl, 1994). This suggests that smaller groups may be more successful in terms of maximizing the amount that each student contributes to the experience, and that a dyad may have the benefit of the most participation and facilitation with the least possibility for loafing, free-riding, interruption or distraction.

The presence of others can also enhance problem solving by prompting evaluation, explanation or reflection; supporting the planning and execution of complex tasks; and contributing new perspectives or novel information or skills that other students may not possess. All these factors could theoretically improve the quality of a group's contribution. For example, others may detect errors and provide immediate feedback to any individual in the group (Schoenfeld, 1989). Groups may also generate a more reflective, explicit or abstract problem representation than a lone individual (Moreland & Levine, 1992, Schwartz, 1995). Therefore, if having additional members of a group increases the frequency of these beneficial behaviors and the likelihood of novel perspectives and skills, then triads should be expected to outperform dyads simply because there are more people in the larger group.

On the other hand, the potential for being evaluated can also have an inhibiting effect and working with others can cause evaluation apprehension, causing poorer performance, and the generation of fewer or less creative ideas (Allport, 1920; Camacho & Paulus, 1995). The potential for critical evaluation exists as soon as a second person is added to the task. If a group does not handle conflict well, then the larger the group, the more likely it is that negative effects may occur. This again suggests an advantage for dyads.

However, while the presence of at least three people in a group may increase the potential for conflict, if this conflict makes the group more likely to think about multiple perspectives and to critically evaluate both "majority" and "minority" stances in terms of evidence and justifications, then positive effects may occur. For this reason it could be predicted that triads may engage in more productive monitoring, evaluation, revision and reasoning than dyads or singletons. To the extent that the presence of a minority position promotes argumentation and evaluation among the members of the group, triads may be a more optimal group size for some learning and problem solving contexts.

In sum, there are some theoretical reasons to suggest that dyads may be an optimal group size for problem solving and learning contexts, and another set of theoretical reasons to suggest that triads may show better performance than dyads. This study provides a direct test of the group size question, whether dyads or triads would perform better on a mathematic-logical reasoning task.

To test this, we used a Letters-to-Numbers modified cryptarithmetic task as first described in Laughlin, Bonner, and Miner (2002). In this task, solvers attempt to decipher a random coding of numbers to letters using logical and mathematical reasoning. Unlike traditional cryptarithmetic problems where all of the information necessary for solution is contained within the problem, the Letters-to-Numbers task requires participants to solicit and assimilate additional information from an external source (the experimenter) through evidence collection and hypothesis testing. Previous studies have shown a benefit for groups in solving these types of problems versus individuals, but have yet to establish the ideal group size for best performance. In this study, we compare triads, dyads, and individual solvers using several measures of problem solving quality and efficiency. Further, we are interested not only in which group size may perform best on this task, but whether heterogeneity in relevant math skills contributes to performance, and if so, which group size might be most likely to take advantage of heterogeneity in background.

Method

Participants

The participants were 90 students at the University of Illinois at Chicago enrolled in an Introductory Psychology course who received course credit for their participation. The participants were randomly assigned to work as either a three-person cooperative group (N=30), a two-person cooperative group (N=30), or as individuals (N=30). Descriptive group information is presented in Table 1.

Gender was matched across the three different group sizes; individuals (m=15, f=15), dyads (m/m=5, f/f=5, m/f=5), and triads (same sex=5 (3 mmm, 2 fff), mixed sex=5 (3 ffm, 2 fmm). No differences were seen in solution rates for males and females in the singletons, or among three types of dyads. Sample sizes in the triads were too small for analysis but showed neither a pattern of increase with either increasing numbers of males or females nor an advantage for same over mixed-sex groups.

All groups in the present analyses were comprised of participants who did not have pre-existing close relationships with one another. Several groups of friends did participate, but their data has been removed.

Procedure

Participants first completed a practice task where they generated as many words as possible using the letters in the words "Lake Michigan" during a five minute period. The main purpose of this task was to act as a warm-up exercise in which all members would start interacting. Such an exercise is important for maximizing performance in ad hoc groups. All groups engaged in this task, with groups averaging around 30 words. There is a significant difference here with triads generating more words than the individuals (triads, M=35.1; dyads, M=28.3; individuals, M = 22.43). More importantly, the task was successful at getting each member of the groups to participate.

After this task, the participants completed a Letters-to-Numbers problem (based on Laughlin, Bonner, & Miner, 2002). The instructions used in this study were taken directly from Laughlin et al (2002). In the Letter-to-Numbers problem, participants are asked to figure out a random mapping of the digits 0-9 to the letters A-J. The objective is to use a series of participant-generated letter equations and hypotheses in conjunction with experimenter feedback to decipher the entire coding of letters to numbers in as few trials as possible. Students need to engage in a reasoning process and use math facts to figure out which letter corresponds to which number.

All participants solved for the same random mapping of digits to letters. At the beginning of each trial, participants were asked to spend some time thinking about the problem and then asked to generate an equation (e.g. A + B = ?). The dyads and triads then discussed the equations until they reached consensus as to which equation to propose. Once the group agreed on an equation, the experimenter told the group the correct answer to the equation in letter form (i.e. A+B=F). Next, each individual crafted a hypothesis (e.g., A=1). The groups discussed individual hypotheses until they reached consensus as to which hypothesis to propose. Once a hypothesis was selected, the experimenter indicated whether the hypothesis was true or false. Finally, the participants had the option of proposing a full coding of the letters to numbers where they received feedback as to whether the entire coding is correct or not, or moving on to the next trial. Participants recorded all equations, hypotheses, and feedback on response sheets to minimize demands on memory.

After completing the Numbers-to-Letters problem (or failing to solve it in 10 trials), several measures of mathematical skill were collected. First, math skill was assessed by having students attempt to solve an algebra word problem within three minutes. The exact problem was "Two trains leave the same station at the same time. Each has enough fuel for a 2000 mile trip. The trains travel in opposite directions. One train travels 60 miles per hour, and the other 100 miles per hour. In how many hours will the trains be 800 miles apart?" The problem used was based on earlier studies in our population (Ash & Wiley, 2002) finding that around a third to a half of our students generally can solve this problem successfully in that time. For the purposes of this project, one measure of math skill was defined as having solved the problem correctly. Second, they attempted an LSAT/Old GRE analytical section logic problem (involving 7 students and various constraints on whose lockers can be adjacent). Finally, the participants completed a survey that asked for demographic information as well as their ACT or SAT math scores, and contained a self-report rating of their math skill on a 1-7 scale, where 7 meant "Not very skilled". After all participants completed the survey, the experimenter explained the purpose of the study, asked if there were any questions, gave them a written debriefing form with references, and thanked them for participating. All sessions were video-recorded. The whole procedure was generally completed in under an hour.

Results

Math Skill

No differences were seen in the average level of math or logical problem solving skill across the three group sizes, all Fs<1. Descriptive statistics on the math skill and other measures for the groups are presented in Table 1.

Only a portion of the students reported an ACT score (N=71), and several reported uncertainty about whether the number they recalled was their math score or the composite. As a result, we felt the data collected on the ability to solve the algebra word problem provided the most complete and reliable data on math skill. Math problem performance did relate to ACT (r=.40, N=71, p<.001) but not self-reported skill (r=-.15, N=89, p<.18). Although effects were not significant, it is important to note that if anything, the ability to solve the math problem was highest in the singleton condition and lower among dyads and triads.

Table 1: Descriptive and Dependent Measures (with SEs) for Each Group Size.

	Singles	Dyads	Triads
N	30	30	30
Groups/Units	30	15	10
Mean Age	19 (1.23)	19.27 (1.39)	18.93 (1.23)
Mean Math ACT	24.1(.79)	24.8(.98)	23.4(.84)
Self-report skill	3.5 (.28)	4.0 (.21)	3.6 (.24)
Proportion solving math problem	0.57(.09)	0.43 (.09)	0.47(.09)
Proportion solving Logic Problem	0.66(.46)	0.53 (.39)	0.63(.45)
Trials to Solution	9.67(.23)	8.87(.41)	6.7 (.56)
Frequency of Repeat Equations	34	11	4
Letters proposed Per Equation	2.10(.02)	2.17(.15)	2.38(.35)
Solutions available Per Trial	1.01(.04)	1.14(.07)	1.64(.18)

Trials to Solution

The main effect of group size was significant, $F(2,52)$= 15.54, MSE=2.13, p<.001. The means and standard errors are displayed in Table 1. Post-hoc tests using LSD indicated that triads reached solution in fewer trials than both individuals (*p*<.001) and dyads (*p*<.002).

Following the nominal group analyses of Laughlin et al (2002), the best, second-best, and third-best individuals were determined by the number of trials to solution for each "group" of three singletons who completed the experiment consecutively. The logic behind this analysis is that a group can perform at the level of the best individual without having to assume any additional advantages of collaboration. This approach revealed that the best individuals in nominal triads had fewer trials to solution (M = 8.5) than the second-best individuals (M = 9.9) and the third-best individuals (M = 10.6). (Participants who did not solve the numbers-to-letters problem in the allotted 10 trials were considered to have required 11 trials.) Similarly, for nominal dyads, the best individuals required fewer trials to solution (M = 8.87) than the second-best individuals (M = 10.47).

Planned contrasts between actual and nominal triad performance indicated that actual triads had significantly fewer trials to solution than the best individuals, $t(18)$ = 2.749, p<.013; second best individuals, $t(18)$ = 4.991, p<.001; and third best individuals, $t(18)$ = 6.49, p<.001 from nominal triads. The comparison to nominal groups is important to show that triad success is not simply due to the higher chance that one member of a group was more intellectually able or skilled, or more likely to "know the answer." Therefore, these results show that there were additional benefits due to collaboration among the individuals in a triad, and that triads were performing better than the highest level of individual performance would predict.

Planned contrasts between actual and nominal dyads indicated that dyads did not have significantly fewer trials to solution than the best individuals, but did have fewer trials to solution than the second best individuals from nominal dyads, $t(28)$ = 3.601, p<.001.

Repetitive Equations

Repetitive equations are equations that have already been offered by the group or individual in either identical or a different form (e.g. A + B = F is the same as F – B = A). The inclusion of repetitive equations can be seen as a failure to monitor or evaluate solution attempts, or as the result of poor reasoning. As can be seen in the frequency data presented in Table 1, singletons produced more repetitive equations than dyads, and dyads produced more than triads. Analyzed as the proportion of singletons (60%), dyads (53%) and triads (20%) to generate repetitive equations, differences in group size approached significance, $X^2(2)$=4.85, N=55, p<.08. This demonstrates that only the triads seem to be engaging in critical evaluation and monitoring of their suggested solutions. Although peers could monitor and evaluate others' suggestions in the dyads, their use of repetitive equations suggests that dyads are not effectively monitoring their performance.

Average Letters Proposed Per Equation

Another measure of the efficiency of problem solving is the number of two-letter equations that are proposed. Proposing only two-letter solutions is an inefficient strategy as it generally will provide less information than an equation with more letters. The longer groups persist in this strategy, the longer it will take them to map all the letters. Although there was a trend of increasing complexity with larger group size as shown in Table 1, a main effect was not seen on this measure, $F(2, 52)$=.939, MSE= .30, p=.397.

Letter Solutions Available Per Trial

In each trial, solvers may discover the solution to a number-letter pair either through experimenter feedback on a hypothesis or through numerical and logical inferences. Using these same processes, group performance can be coded and scored to determine the total number-letter pairings that could be identified on any given trial. In this way, the quality of the proposed equations can be evaluated. Those groups that have a higher average number of letter solutions available per trial are choosing more informative equations. A significant main effect of group size was found for this variable $F(2,52)$=15.33, MSE=.10, p < .001. Post hoc tests indicated that triads had more letter solutions available per trial than dyads and singletons. Triads also outperformed the best (M = 1.18), second-best (M = .97), and third-best individuals (M = .89) on this measure. This can be taken as evidence of superior reasoning in proposing solutions among the triads, and demonstrates a specific strategic advantage due to collaboration among triads.

Effect of Heterogeneity of Groups

To explore group performance as a function of heterogeneity in background skills, the number of members of the group with high math skill was examined. For the purposes of this analysis, high math skill was defined as having solved the algebra word problem correctly. The pattern of results is presented in Figure 1. First, no significant differences were seen between high (N=17) and low (N=13) math skill individuals in the singleton condition, t(29)=1.58, p<.12. Although there does seem to be a trend toward skill improving performance, the improvement is no where near the level observed some dyads and triads. This is important as it indicates that possession of math skill by a single individual alone does not lead to the most effective level of performance on this task.

A 2x3 ANOVA (dyad vs. triad; zero, one and two high math members) was performed on number of trials to solution. While both main effects for group size, F(2,19)=3.01, MSE=2.17, p<.07 and number of math members, F(1,19)=2.26, p<.15 approached significance, a significant interaction was obtained, F (2,19)=3.62, p<.05. When two members of a dyad (2 groups) or triad (5 groups) had high math skill, either group was more successful in problem solving than when no members had math skill. However, only triads (4 groups) were able to take advantage of having a single high math member. Dyads with a single high math member (9 groups) did not perform differently than dyads with no high math members (4 groups) or singletons.

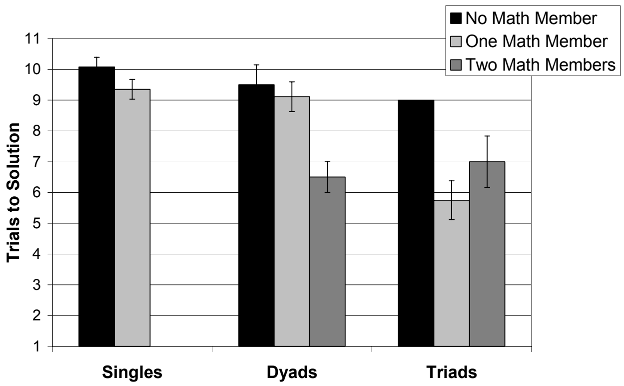

Figure 1: Heterogeneity of Math Background, Group Size and Trials to Solution

The one triad with all low math members also performed similarly to singletons. Although there is an uneven distribution of groups to condition, the striking effect here is that triads may be better able to take advantage of group heterogeneity when only one member has skills that are needed to solve a problem. Perhaps the most important comparison to note is between individuals with math skill, the dyads with a single math member, and the triads with a single math member. Here the number of skilled members is held constant at one, but only in the triad context is there an improvement in problem solving performance.

Discussion

The results of the present study show that on a modified cryptarithmetic problem solving task, triads outperform dyads, individuals and even the best individuals in a nominal group analysis. Triads solved problems in fewer trials, were more effective in their equation selection, were better at monitoring the redundancy of their solution attempts, and were better able to take advantage of group heterogeneity on this task.

The nominal groups analysis showed that triads did not do better just due to the chance of having a member who "knew the answer". The group performed better than the best individuals, showing that there was something specific to the collaboration in these groups that was responsible for the high level of problem solving. Another way of seeing this is to note that not all group types experienced superior problem solving. Dyads and triads with no math members were not better than singles, nor were dyads with a single math member. So, increasing the number of members did not in and of itself improve performance. Similarly, having a member with math skill also did not guarantee better performance. Instead, we observed an interaction between the size of the group, and the distribution of math skill. In dyads, both members had to have math skill to be effective, while in triads, as long as one member had math skill, the group was more effective.

A very interesting question that we are still pursuing with this data is a more detailed analysis of how successful collaboration and coordination was achieved in our mixed-knowledge triads. What is clear from the data so far is that triads are more likely to effectively monitor each others suggestions, as shown in the lower rate of redundant equations. They also seem to engage in more effective reasoning, as they propose more informative equations. These results suggest that the presence of the third person in a triad may have offered some affordances over and above the presence of another peer in a dyad that supported better evaluation, reasoning, and the contribution of otherwise unshared information among the participants.

Our next step is to analyze the group discussions to better understand the dynamics in effective groups. Through protocol analysis, Barron (2003) reported that responsiveness among group members was critical for collaborative learning benefits among students who learned to solve arithmetic problems in triads. We will be very interested in how successful and less successful groups in our sample interact, which strategies they use, what knowledge is shared, and what kinds of challenges and explanations are offered as our students attempt to solve these problems.

There are some other important future directions for this research. First, the present results were found with ad hoc groups. It is an open question whether similar or different results will be seen with students who work together in teams for an extended amount of time or who are otherwise more familiar with each other (such as being classmates). This step is necessary to help determine which group size might be most effective in a classroom context.

Further, the present set of results shows advantages of collaboration only in group-level performance. A key question is whether these advantages will be seen when individuals need to use what they learned in this context on later transfer tasks where they solve problems alone. Both of these issues represent the next steps in our research program, seeing if the advantages observed here will also occur in intact teams with classes, and on individual transfer measures.

When are three heads better than two? The present study suggests that triads are a better group size than dyads for performing this reasoning task, particularly when triads are heterogeneous in relevant skills for the task. Several potential benefits of small group problem solving occurred most clearly in this context. Importantly, when math skill was held constant with one high math member across individuals, dyads and triads, only triads showed an advantage in their problem solving. Overall, triads engaged in better monitoring, evaluation and higher quality reasoning processes in proposing solution attempts. The exact mechanisms and interaction patterns responsible for these effects are still under investigation, but the results of this study suggest that gains from cooperative collaboration were most likely to be realized in triads.

Acknowledgments

This research is funded by a CAREER grant to Jennifer Wiley from the National Science Foundation. All opinions expressed are those of the authors and do not reflect those of the funding agency.

References

Allport, F. H. (1920). The influence of the group upon association and thought. *Journal of Experimental Psychology, 3*, 159-182.

Ash, I. & Wiley, J. (2002). A-ha! I knew it all along. Hindsight bias in insight and algebra problems. *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 63-69). Hillsdale, NJ: Erlbaum.

Barron, B. (2003). When smart groups fail. *The Journal of the Learning Sciences, 12*, 307-359.

Camacho, L. M., & Paulus, P. B. (1995). The role of social anxiousness in group brainstorming. *Journal of Personality & Social Psychology, 68*, 1071-1080.

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward & S. M. Smith, (Eds), *Creative thought: An investigation of conceptual structures and processes.* Washington, DC, US: American Psychological Association.

Dugosh, K. L., Paulus, P. B., Roland, E. J., & Yang, H. (2000). Cognitive stimulation in brainstorming. *Journal of Personality & Social Psychology, 79*, 722-735.

Hill, G. W. (1982). Group versus individual performance: Are N + 1 heads better than one? *Psychological Bulletin, 91*, 517-539.

Hoffman, L., Harburg, E., & Maier, N. (1962). Differences and disagreement as factors in creative group problem solving. *Journal of Abnormal and Social Psychology, 64*, 206–214.

Laughlin, P. R., Bonner, B. L., & Miner, A. G. (2002). Groups perform better than the best individuals on letters-to-numbers problems. *Organizational Behavior and Human Decision Processes, 88*, 605-620.

Maier, N. R. F. (1967). Assets and liabilities in group problem solving: The need for an integrative function. *Psychological Review, 74*, 239-249.

Moreland, R. L., & Levine, J. M. (1992). Problem identification by groups. In S. Worchel, W. Wood, & J. A. Simpson, (Eds.), *Group processes and productivity.* Newbury Park, CA: Sage.

Nemeth, C. (1986). Differential contributions of majority and minority influence. *Psychological Review, 93*, 23–32.

Paulus, P. B. & Yang, H. (2000). Idea generation in groups: A basis for creativity in organizations. *Organizational Behavior & Human Decision Processes, 82*, 76-87.

Salomon, G. & Globerson, T. (1989). When teams do not function the way they ought to. *International Journal of Educational Research, 13*, 89-99.

Schoenfeld, A. H. (1989). Ideas in the air: Speculations on small group learning, environmental and cultural influences on cognition and epistemology. *International Journal of Education, 31*, 71-88.

Schwartz, D. L. (1995). The emergence of abstract representations in dyad problem solving. *Journal of the Learning Sciences, 4*, 321-354.

Stasser, G., Stewart, D., & Wittenbaum, G. (1995). Expert roles and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology, 57*, 244–265.

Steiner, I. D. (1972). *Group processes and productivity.* New York: Academic Press.

Stroebe, W. & Diehl, M. (1994). Why groups are less effective than their members: On productivity losses in idea-generating groups. *European Review of Social Psychology, 5*, 271-303.

Taylor, D. W., Berry, P. C., & Block, C. H. (1958). Does group participation when using brain storming facilitate or inhibit creative thinking? *Administrative Science Quarterly, 3*, 23-47.

Wiley, J. (1998). Expertise as mental set: Negative effects of domain knowledge on creative problem solving. *Memory & Cognition, 26*, 716-730.

Wiley, J. & Bailey, J. (2006). Effects of collaboration and argumentation on learning from web pages. In A.M. O'Donnell, C. E. Hmelo-Silver & G. Erkens (Eds.) *Collaborative Learning, Reasoning, And Technology*. Hillsdale, NJ: Erlbaum.

Wiley, J. & Jolly, C. (2003). When two heads are better than one expert. *Proceedings of the Twenty-Fifth Annual Conference of the Cognitive Science Society.* Hillsdale, NJ: Erlbaum.